"一带一路"智库报告主题挖掘与演化研究 *

祁瑞华 [1,2] 付豪 [2]

[1] 大连外国语大学语言智能研究中心 大连 116044
[2] 大连外国语大学软件学院 大连 116044

摘要：[目的/意义]作为国家"软实力"和"话语权"的重要组成部分，智库对政府决策、企业发展、社会舆论与公共知识传播等方面具有深刻影响。在对"一带一路"合作倡议的认知上，智库以及智库专家的研究成果，在对外合作、舆论导向、项目评估等方面发挥着重要作用，同时也能够反映"一带一路"合作倡议的工作重点等信息。本文通过量化研究中国智库对"一带一路"倡议研究的主题分布与演化历程，得到"一带一路"倡议在中国智库视角下的重点关注、研究方向和研究发展脉络，为"一带一路"倡议相关研究的推进提供整理与参考。[方法/过程]本文结合 LDA2vec 主题模型与词向量语义相似度计算方法，对于《全球智库报告 2020》榜单中的中国智库，在 2013—2020 年间关于"一带一路"倡议研究的 3,052 份报告进行主题挖掘以及主题演化研究，并进行可视化呈现，分析主题内容与演化特征。[结果/结论]我国智库对"一带一路"倡议的研究热点主题逐年覆盖政治、经济、文化、社会、军事、环境等领域，且研究热点有着鲜明的基调，如"互利互信""互联互通""命运共同体""共赢"等，主题演变的过程体现出明显的主题继承性、主题融合性等特征，这也体现出"一带一路"倡议逐渐全面的发展过程，也能够说明国内智库对"一带一路"倡议的研究重点具有持续性和拓展性。

关键词："一带一路" 主题挖掘 主题演化
分类号：G251.2；D812
DOI: 10.19318/j.cnki.issn.2096-1634.2022.05.02
开放科学（资源服务）标识码（OSID）

1 引言

作为国家"软实力"和"话语权"的重要组成部分，智库的研究文本具有内容权威、信息密度大、学科领域融合等特点，也更能体现研究对象的研究价值。研究智库文本有利于进一步探索情报学发展方向，同时对情报学在跨学科领域的研究起到重要的推进作用。

2013 年，中国国家主席习近平提出国家顶层合作倡议——"一带一路"合作倡议，该倡议一经提出就深受国际瞩目，以其合作特色和发展优势，吸引越来越多的国家研究和参与。在研究层面，智库的研究重点和领域能够体现出"一带一路"倡议的整体发展和着力点，在"一带一路"倡议的认知和推动上，智库研究成果发挥着不可或缺的作用。研究智库对某一国际事务的观点和看法，总结分析智库研究的重点问题，正在成为研究国际关系发展与变化的重要手段，因此，分析智库研究成果能够为"一带一路"体系建设和未来深化合作水平提供合理建议与思路。

本文根据美国宾夕法尼亚大学"智库与公民社会项目"（Think Tanks and Civil Societies Program，TTCSP）发布的《全球智库报告 2020》[1] 中上榜的中国智库以及其关于"一带一路"倡议的研究报告为研究对象，运用主题挖掘和主题演

* 本文系大连外国语大学计算语言学与人工智能创新团队（项目编号：2016CXTD06）研究成果之一。
作者简介：祁瑞华，大连外国语大学软件学院教授，硕士生导师，博士，E-mail：rhqi@dlufl.edu.cn；付豪，大连外国语大学软件学院图书情报硕士研究生，E-mail：2990849010@qq.com。

化相关文本信息处理技术，得到“一带一路”倡议在我国的研究现状、研究发展脉络以及研究主题间的关系等，为我国推进“一带一路”倡议有关建设、探索情报学视角下的“一带一路”倡议等研究，提供一定借鉴。

2 研究现状

在对“一带一路”倡议的智库认知和关注重点的研究上，申钟秀选取 9 家印度主流智库的研究报告作为研究样本，分析其对“数字丝绸之路”建设的认知、影响因素等，分析得到我国“数字丝绸之路”加强与印度合作方面的建议 [2]；姜昱同等对 9 家美国智库发表的 16 篇研究文献进行观点研究，分析美国智库对“健康丝绸之路”的关注重点，针对其负面评价提出应对意见 [3]；王飞等以拉美地区智库作为研究分析主体，得到拉美国家对“一带一路”的认知和参与方式，并对“一带一路”倡议下中国与拉美地区的现状、问题等进行总结 [4]；干保柱等从日本智库的“一带一路”倡议研究入手，着重对相关评价进行分析，得到日本各界对“一带一路”倡议的认知情况与转变历程，为促进中日合作提供更多切入点 [5]；陈媛通过对荷兰国际关系研究所“一带一路”相关文献的研究，分析总结其态度和成因，并对“一带一路”的传播策略提出建议 [6]；万盖利 • 阿纳斯塔斯（Vangeli Anastas）等运用思想传播理念，研究中国与中东欧智库在社会与政策层面的相互影响，并提出了一个基于传播反作用于社会的理论框架 [7]，等等。

在智库领域的主题挖掘与主题演化研究上，齐欣以推特平台上美国智库专家的对华分析数据作为研究对象，基于 LDA 主题模型分析其议题以及热度变化，探究美国对华观察视角 [8]，同时其也基于主题挖掘和其他计量方法，对美国智库专家的推特涉华信息获取途径进行探究，分析信息源的出处和获取入口 [9]；王晰巍等基于 LDA 主题挖掘和社会网络分析法构建了突发事件下网民情感演化模型，追踪舆情阶段的主题演化与情感演化 [10]；孙瑞英等采取多学者层次模型，对高校智库研究文献不同时期的研究主题进行对比分析，动态把握主题演化规律 [11]；王扶东等基于国内“一带一路”倡议的相关科技文献，结合 LDA 主题模型进行主题挖掘，并从关键词共现角度进行主题交叉演化研究 [12]；胡潜等运用用户和主题关联矩阵研究方法，建立虚拟社区的智能推荐方法，兼顾了数据挖掘和信息推荐方法 [13]。

综上所述，对“一带一路”倡议智库领域的研究主要为海外智库对“一带一路”倡议的认知研究，总结性和定性分析较多。从“一带一路”倡议研究出发，结合主题挖掘和演化角度对智库领域的相关研究较少，智库领域的主题挖掘和演化研究以 LDA 主题模型和关键词共现方法 [14] 为主，结合网络化、图谱化下的主题关联与演化研究 [15] 也有较好的应用，但是涉及更加精确、更细粒化的定量分析对“一带一路”倡议智库研究较少，在更深层次、更精准化等方面的探索应当进一步提升，尤其是深度学习模型 [16] 与算法的使用。本文基于机器学习和深度学习的文本信息处理技术，结合 LDA2vec 主题模型和语义相似度计算，对“一带一路”倡议智库报告主题及其演化进行研究，创新性地采用词向量和语义维度等研究视角，为相关研究提供一定的探索经验。

3 基于 LDA2vec 的“一带一路”智库主题挖掘与分析

3.1　LDA2vec 主题模型

LDA2vec 主题模型是克里斯 • 穆迪（Chris Moody）等于 2016 年提出的一种以深度学习为基础的主题模型，其主要思想是将 LDA 主题模型和 word2vec 词向量模型相结合进行主题衡量。当前，LDA 主题模型的应用范围最广，对解释性强的文档主题提取效果较好；Word 2vec 表示方法通过 Skip-gram 和 CBOW 模型对词向量进行训练，被广泛应用于对文本表示效果的改进。LDA2vec 主题模型结合 LDA 与 Word2vec 两种方法的优势，能够实现词向量、主题向量与文档向量的优化组合，深度学习理念中神经网络的引入也使得 LDA-2vec 更加精确地体现文档主题的聚类与主题词的关系。目前，LDA2vec 的应用领域正在不断地拓宽，模型改进也处于不断发展之中。

LDA2vec 主题模型主要思想如图 1 所示，是利用文本内容主题概率分布来学习向量嵌入，通过向量的匹配与计算，得到主题的提取结果。

LDA2vec 主题模型主要依靠上下文向量来进行主题的预测和抓取，上下文向量的生成以词向量和文档向量作为主要核心内容，词向量是借助 Skip-gram 模型输入单词，Skip-gram 作为神经网络能够预测出输入词的上下文单词，通过 INPUT 层输入词 $W(t)$，经过 projection 层的神经网络实现向量空间映射，OUTPUT 层输出对上下文的词预测，即 $W(t+1)$、$W(t-1)$ 等。文档向量是融合了文档权重向量和主题矩阵的集合体，文档权重向量表示的是文档中每个主题的权重，主题矩阵是每个主题的对应向量表示。文档向量和单词向量的协同生成提供了每个单词的上下文向量。那么 LDA-2vec 便能够在词嵌入和上下文向量组合的嵌入训练的同时，对主题和文档进行表征，模型训练结果也能够提高主题提取的准确性以及数据利用的多样性。

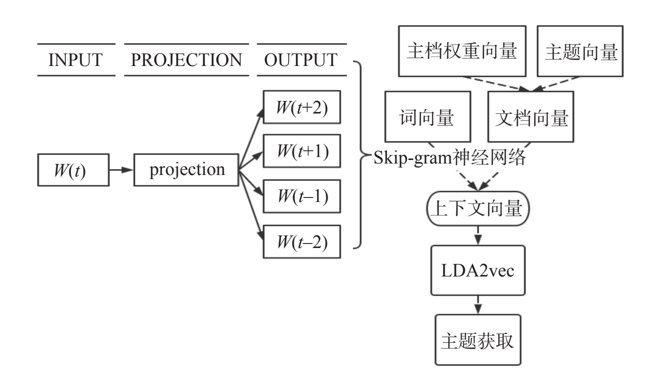


图 1 LDA2vec 主题模型流程图
Figure 1 Flow figure of topic model LDA2vec

3.2 基于 LDA2vec 的“一带一路”智库主题挖掘

3.2.1 数据来源 本文选择《全球智库报告 2020》上榜的 8 家中国智库，在这 8 家智库的官方网站对研究报告、专家观点、咨询分析等多种研究信息进行爬取，检索词为“一带一路”“Belt and Road”“B&R”，检索时间段为 2013—2020 年，包括分析报告、专家观点、资讯评价、专题点评等在内的检索结果均被认定为“一带一路”智库研究成果，将爬取结果统一保存为 CSV 文件，存入本文建立的“一带一路”智库研究报告数据库中。本文获取中国智库研究报告共 3,239 份，经过分类整理和人工筛选等方法去除无关、重复数据后，得到研究样本报告共 3,052 份，研究文本字数共 312,223 字，数据来源如表 1 所示。

表 1 《全球智库报告 2020》中国智库研究数据统计表
Table 1 Statistical research data table of Chinese think tanks in *Global Think Tank Report 2020*

序号	智库名称	智库成立时间	“一带一路”样本	文本字数
1	中国现代国际关系研究院	1980 年	30	42,702
2	中国社会科学院	1977 年	1,008	69,927
3	清华 - 卡内基全球政策中心	2010 年	128	10,989
4	国务院发展研究中心	1985 年	930	72,332
5	中国国际问题研究院	2014 年	74	7,050
6	全球化智库	2008 年	225	39,858
7	北京大学国际战略研究院	2013 年	66	11,250
8	上海国际问题研究院	1960 年	591	58,115

3.2.2 文本预处理 由于初步整合的文本数据较为混杂无序、数据量大，无法直接用于 LDA2vec 主题模型中，因此，需要对数据进行包括切词分词、建立停用词词典、去停用词、词性标注等文本预处理操作。本文以中文文本为主，使用 Python 中 jieba 中文分词库实现分词。去停用词处理是对文本数据自动过滤无意义词汇的操作，是提升数据利用效率的重要方式，能够提高主题模型的性能和结果的准确性。建立停用词词典是去停用词操作的基础，也称停用词表，是筛选剔除词汇、保留词汇的依据。停用词表的建立一般需要基础词表和专业词表组合使用，本文选择 Github 上的中文停用词表作为基础停用词表，专业停用词表根据实验中模型迭代不断更新完善。

停用词处理后的文本滤掉了大量无意义词汇，通过词性标注可以挑选出信息密度较高的词汇，便于研究，词性标注是按照词的特点对文本数据进行类别划分的方法，主要根据语法特点开展，词性分类有两大类：一类是实词，包括名词、动词、代词、形容词等；另一类是虚词，包括介词、连词、助词等。本文主要对实词进行整合，进行下一步分析。

3.2.3 主题提取结果与分析 文本经过预处理后，按照 2013—2020 年的时间排序，运用 LDA2vec 模型迭代训练，得到我国智库研究报告的主题与主题词。根据多轮实验结果和主题相关性考量，以及模型的困惑度（perplexity）最低值来确定每一年文本的主题数，并确定主题词统一选择为 15 个最相关的主题词，模型迭代次数为 25 次效果最好。最后得到的主题提取结果如表 2 所示。

表 2 中国智库“一带一路”研究主题提取结果（部分）
Table 2 Subject extraction results (part) of Chinese think tanks’ “ Belt and Road Initiative” study

年份	主题	主题词（部分）					
2013 年	亚洲国家经济合作	亚洲	贸易	稳定	发展	合作	经济
2014 年	基础设施建设与贸易、投资	基础设施	政策	贸易	共同体	投资	产业
	沿线国家政治合作	政治	利益	合作	丝绸之路	区域	沿线
	文化交流与社会合作	文化	合作	沿线	历史	社会	关系
2015 年	国内就业创业和人才培育	中西部	机遇	创业	培育	全球化	就业
	务实合作繁荣共赢伙伴关系	丝绸之路	务实	繁荣	经济带	共赢	互利互信
	金融监管投资与电子商务	监管	融资	业务	成本	资产	电子商务
	国际势力争端与反对恐怖主义	领土	冲突	争端	独立	反对	恐怖主义
	沿线地区军事冲突	战争	领土	势力	中亚地区	军事	冲突
2016 年	制造业与金融市场运营	金融机构	投资	租赁	运营	便利化	制造业
	对外开放互惠互利合作	对外开放	自贸区	丝绸之路	共赢	互利	合作
	学术交流与人文合作	交流	学术	务实	大学	丝绸之路	合作
	军事争端与反恐怖主义	恐怖主义	欧盟	谈判	军事	地缘	争端
2017 年	平等互利合作新格局构建	新格局	共享	互鉴	共同体	平等	命运
	金融危机与贸易竞争以及冲突	金融危机	风险	防范	矛盾	冲突	竞争
	电子信息商务行业和物流流通	金融服务	信息主体	金融	融资	资金	网络
	天然气、油气、海洋能源贸易	天然气	油气	产业链	进出口	资源	海洋
	贸易节点统筹安排与互联互通	经济带	节点	互联互通	网络	统筹	打造
	政治合作与贸易协议谈判	保护主义	谈判	合作	双边	协定	政府
2018 年	命运共同体和高质量对外开放	命运共同体	高质量发展	共享	互联互通	改革开放	治理
	政治联盟和外交政策	联盟	外交	合作	政治	政府	对外
	绿色治理理念和经济发展	高质量	环境	协调	发展	绿色	体系
	科技技术创新发展与贸易	科技产品	技术	优势	数字	创新	贸易
	基础设施建设与互联互通	非洲	拉美	基础设施	建设	互联互通	交流
2019 年	基础设施与市场投资	治理	基础设施	市场	投资	开放	贸易
	文化产业发展与高校建设	产业	建设	高校	政策	社会	文化
	产业转移与市场贸易发展	转移	产业	贸易	企业	基础设施	发展
	绿色发展与金融体系服务	发展	经济	体系	金融	开放	企业
2020 年	高质量经济与基础设施建设	高质量	互联互通	治理	经济建设	体系	基础设施
	互联互通命运共同体全球治理	共同体	治理	全球化	丝绸之路	互联互通	多边

由主题提取结果可知，中国智库对“一带一路”倡议的研究主题呈现先丰富后融合的趋势，可以总体分为两个阶段：第一个阶段，是2013—2017年，研究主题逐渐增多和拓展，这体现了研究的不断发展；第二个阶段，是 2017—2020 年，研究主题数虽然减少，但是研究主题的全面性和融合性上不断提升，研究主题的综合性变强，这表明研究的成熟和稳定。

第一个阶段是从 2013 年“一带一路”倡议的提出开始，着重研究“一带一路”的地区性经济合作以及影响，主要是在亚洲地区的贸易往来与经济合作，注重稳定发展的理念，集中在中国同俄罗斯、印度等国家的贸易往来研究，反映了“一带一路”初期经济合作处于起步阶段。2014 年，研究热点开始增多，除了经济贸易合作以外，政治、社会等领域的研究得到拓展，同时基础设施建设

成为“一带一路”倡议在经济领域的重要研究主题，也反映了“一带一路”倡议开始推进具体措施。2015 年，在互利互信、务实合作的主基调下，研究热点为贸易伙伴关系，以及就业创业、金融管理、人才引进等，突出“一带一路”倡议对国内发展的影响，其他研究主题同国际形势结合紧密，涉及中亚地区的军事冲突、领土争端、国内局势、反对恐怖主义等，同时美日同盟、俄罗斯与乌克兰冲突等研究热点也具有时代特色。2016 年的智库研究持续在经济、军事、政治领域深化，突出制造业和金融业的发展研究，深化对外开放与合作，体现“一带一路”倡议国际影响力的扩大。值得注意的是，新增了文化领域的研究，包括学术交流与合作、人文对话、大学之间的对接与交流等，体现了“一带一路”倡议合作领域的扩展。2017 年，经济领域的研究热点内容较多，以平等合作、能源贸易、电子商务和物资流通为主，围绕“一带一路”沿线亚欧地区领导人贸易协议的签署进行分析，强调应对保护主义这一挑战。同样，分析风险挑战的主题包括金融危机应对、贸易冲突应对、贸易谈判等，此类风险挑战分析主题的出现可能与“一带一路”倡议的推进遇到一定的阻碍有关，在 2017 年表现得尤为显著，风险性研究的出现加上之前的研究可能是 2017 年研究热点较为丰富的原因。

第二个阶段中，2018 年的研究热点分布在政治领域的外交关系和经济领域的基础设施建设、科技创新发展上，另外还在环境治理领域发力，绿色治理研究主题的出现说明研究向高质量推进的方向发展，注重科学协调发展。2019 年的研究主题继续在基础设施建设、绿色发展上深化，并且文化领域重新成为研究热点，主题为文化产业的发展，这表明文化合作上的推进有成效，经济领域的研究主题出现了产业转移的字眼，与“一带一路”倡议带来的国家产业变化有关，能够体现“一带一路”倡议对国家发展的深刻影响。2020 年的研究热点，像高质量经济发展、全球治理和互联互通，更多的是融合之前研究的总结，是“一带一路”倡议的整体影响和成果体现。2020 年的基础设施建设的研究热度仍然不减，这一研究点不断得到深化研究，也说明我国在基础设施建设方面的突出优势。

综合来看，我国智库对“一带一路”倡议的研究领域呈现逐年丰富的趋势，并且时代特征明显，从经济领域的研究开始，逐渐扩展到政治领域，发展到军事、社会、文化、科技等领域，而像军事领域主题的出现，也跟当时的国际形势有密切关联。我国智库的研究热点也同“一带一路”倡议的内涵和重要举措密切相关。例如，当人类命运共同体这一主旨出现的时候，研究重点也都带有合作、交流等主题；互联互通这一主旨出现时，也伴随着出现互利互信、统筹等研究主题。

4 “一带一路”智库研究主题演化分析

4.1 语义相似度计算

语义相似度计算能够从较深层次的语义维度发现语句之间的关联，脱离了单纯的表层文本衡量。语义相似度计算通过在向量空间对文字的表征，使得文字含义之间的发现和关联更加精确。在主题演化的分析中引入语义维度，同基础的时间维度相结合，将向量空间的语义信息带入主题演化研究中 [12]，能够更准确地呈现主题演化的脉络，也能从更细粒化角度发现主题间的关联，分析主题演化趋势。

余弦相似度计算是基于向量的余弦值计算语义相似度的方法，目前在模型适配性上表现较好。采用余弦相似度算法进行语义维度的主题演化研究时，各个主题的相似度数据可以衡量主题的演化关系。例如，主题融合、主题分裂和主题延续等关系，是本文结合时间维度和语义维度进行主题演化研究的重要方面。

余弦相似度计算的是向量空间中两个单词的方向差异，即余弦值，从而得到相似度。夹角越小，余弦值越接近 1，则表示两个向量越相似；反之亦然。因此，余弦值的取值在 –1 到 1 之间。余弦相似度的计算公式如公式（1）所示，其中 ***A***，***B*** 均指代的是计算的向量集合。

$$cos(\theta)=\frac{A\cdot B}{\|A\|\cdot\|B\|}=\frac{\sum_{k=1}^{n}a_k b_k}{\sqrt{\sum_{k=1}^{n}a_k^2}\cdot\sqrt{\sum_{k=1}^{n}b_k^2}} \quad (1)$$

对于计算结果常出现负值，难以直观进行比较的问题，本文选择对余弦相似度计算结果进行

改进，引入余弦距离（Cosine_distance）的定义，即用 1 减去获得的余弦相似度（Cosine_similarity），如公式（2）所示。

Cosine_distance=1–Cosine_similarity　　（2）

余弦距离可以解决计算结果的负值问题，能够将相似度计算结果统一值域，即 [0，2] 的范围。为了便于研究，本文将值域限定为 [0，1]，用 Cosine_similarity/2 表示的取值范围为 [0，1]，且计算结果数值越小，余弦距离越小，表明计算样本之间越相似，当完全相同时，余弦距离为 0，反之为 1。

4.2　主题演化趋势可视化分析

按照语义相似度计算结果与细化指标，总结出各年主题演化关系，并利用 Pyecharts 工具进行可视化。为了便于展示，本文将整体演化图像拆分为两个部分，如图 2 和图 3 所示。其中，图 2 表示 2013—2017 年的主题演化趋势图，图 3 表示 2017—2020 年演化趋势图。

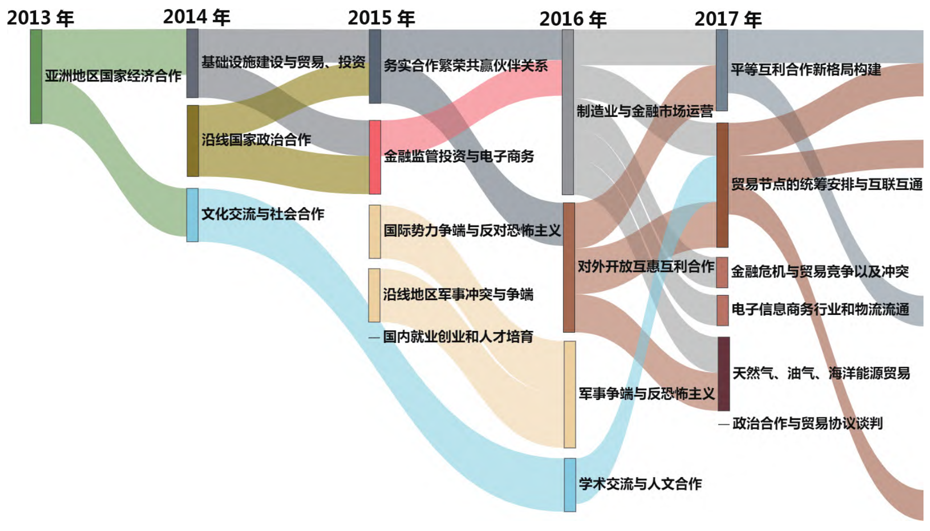


图 2　中国智库“一带一路”研究主题演化趋势图（2013—2017）

Figure 2　Theme evolution trend figure of Chinese think tanks’ “Belt and Road Initiative” study (2013–2017)

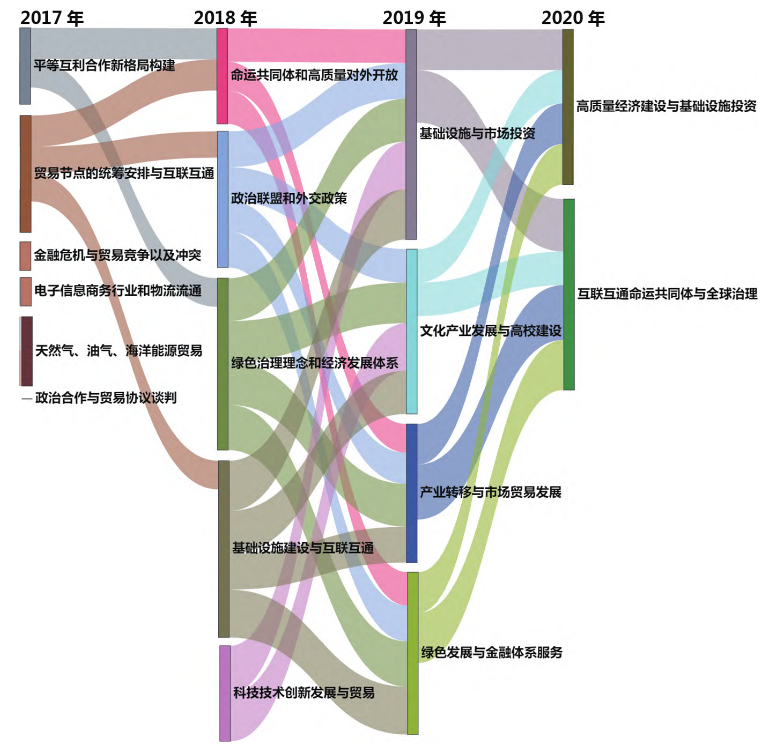


图 3　中国智库“一带一路”研究主题演化趋势图（2017—2020）

Figure 3　Theme evolution trend figure of Chinese think tanks’ “Belt and Road Initiative” study (2017–2020)

由图 2 和图 3 可以看出，结合了语义维度的我国智库“一带一路”研究主题演化趋势具有较为明显的阶段性。

第一个阶段为 2013—2017 年，主题演化的规模较小，发生演化的领域有经济、政治、文化和军事。2013—2016 年主题演化呈现逐渐丰富的特征，研究领域呈现拓展，至 2017 年研究主题虽多但发生演化的主题较少，主题演化规模有所减小。这表明我国智库对“一带一路”的研究在不断发展，对应着“一带一路”的影响领域也不断增多。

在 2013—2017 年的主题演化中，2013 年从经济领域研究出发，以亚洲地区的经济合作为研究基础。2014 年由 2013 年的研究主题分裂出两个方向：一个方向是基础设施建设与投资，是对经济领域研究的深化，也是我国“一带一路”倡议的主要优势；另一个方向是文化交流与社会合作，从经济合作延伸到文化合作，标志着研究领域的拓展和覆盖层次的增多。2015 年的研究主题在政治和经济领域由 2014 年的研究主题融合演化而来，包括政治合作务实伙伴关系、经济领域金融行业的监管。2015 年拓展出了军事领域的研究主题，主要是“一带一路”沿线国家的国家安全与反恐举措以及领土争端和军事冲突，融合形成了 2016 年的军事领域研究，即“一带一路”沿线国家的军事争端与反对恐怖主义相关主题。2016 年其他研究主题部分继承了 2014 和 2015 年的研究，包括文化领域的学术交流与人文合作、对外开放互惠互利等研究主题。经济领域在金融问题上继续研究，同政治合作结合，对“一带一路”相关国家制造业的发展与合作进行分析。基于 2016 年的制造业与金融市场运营和对外开放合作这两个研究主题，2017 年的研究主题出现了平等互利新格局的构建、金融危机和贸易竞争、电子商务与物流以及能源贸易等，“一带一路”贸易节点的互联互通这一研究主题还融合了 2016 年学术交流与人文合作这一研究主题，表明对“一带一路”贸易枢纽的建设的相关研究涵盖的内容更加全面。

第二个阶段是 2018—2020 年，研究领域趋于稳定，主题演化丰富，标志着我国智库对“一带一路”的研究趋于成熟，并且在不同领域持续深入，也说明“一带一路”倡议的推进处于稳定并且深化的阶段。2018 年与 2019 年的主题演化程度最高，2019 年与 2020 年的主题演化也十分丰富，说明在 2018—2020 年时期我国智库研究方向上一致，研究热点主题具有稳定而持续的特征。

在 2018—2020 年的演化阶段中，2018 年的研究主题主要由 2017 年的互利合作和贸易统筹演化而来，包括高质量对外开放、政治联盟、经济发展和基础设施建设研究，并在绿色发展体系、科技创新发展等维度拓展。2019 年与 2018 年的研究主题相关性较高，研究主题均由 2018 年政治外交、绿色发展和基础设施主题演化而来。2019 年的文化以及经济领域研究主题还融合了 2018 年的科技创新研究，表明“一带一路”倡议基础设施建设同科技发展的协同发展。此外，2019 年的经济领域研究主题，如基础设施、市场贸易和金融体系等，均为 2018 年高质量对外开放研究主题的继承。2020 年的研究主题可以概括为高质量经济发展与人类命运共同体的建设，涵盖了政治、经济、文化等领域的研究，全部是由 2019 年研究主题的继承融合而来，是 2018 年、2019 年研究的延续。这一时间段的主题演化多体现出继承性和融合性，研究的持续性和稳定性较好，标志着研究趋于成熟。

总体来看，中国智库在 2013—2020 年的“一带一路”倡议研究主题演化趋势呈现逐渐丰富、持续发展的特征，“一带一路”倡议智库研究一直持续发展和深入，且研究领域呈现先拓展丰富，后深入细化的特点。在演化过程中，研究领域下的热点主题逐年多样，主题演化的主要领域集中在政治、经济、文化领域，主题演化趋势重点领域同主题提取结果反映出的热点主题领域一致。由此可见，“一带一路”倡议的实际覆盖领域不断拓展，影响层面逐渐增多，“一带一路”合作倡议的成效显著。

5 总结与展望

本文基于 3,057 份中国智库“一带一路”倡议研究报告，结合主题挖掘、主题演化理论与方法，对其研究热点进行整理归纳和演化分析。本文创新性在于从主题挖掘与演化分析出发，引入基于深度学习思想的 LDA2vec 模型和基于词向量的文本相似度计算方法，以更细化、量化的方法，结合“一带一路”倡议实际，对我国智库视角下研

究热点和发展脉络进行梳理。

研究发现，我国智库对“一带一路”倡议研究的主题鲜明，每年的研究主题均有各自特点，研究领域逐年覆盖政治、经济、文化、科技、军事、社会等领域，突出强调合作、互利、命运共同体、共赢、基础设施建设等，对我国的制造业、金融行业、文化行业等治理有明显的侧重。各年的研究主题之间均有关联，互有主题演变过程中的主题分裂、主题演化、主题继承、主题创新等过程，主题关系密切，这也说明我国智库在研究“一带一路”倡议过程中的相互借鉴与学科融合，主题发展过程波动性较小，主要研究领域稳定，主题演化具有阶段性。

在未来的研究中，将会引入主题强度、主题情感等更多语义维度下的研究，也将在智库特点及其相关文本特征研究上进一步深化，探索更多的研究层面。

参考文献：

[1] MCGANN J G. 2020 Global go to think tank index report[R/OL]. [2022-03-14]. https://repository.upenn.edu/think_tanks/18/.

[2] 申钟秀 . 印度智库对“数字丝绸之路”的认知评析 [J]. 情报杂志 , 2021, 40(6): 111-118, 207.
SHEN Z X. Some comments on Indian think tanks' perception of the “Digital Silk Road”[J]. Journal of Intelligence, 2021, 40(6): 111-118, 207.

[3] 姜昱同 , 冯卫国 . 美国智库对“健康丝绸之路”倡议的关注重点与中国应对 [J]. 情报杂志 , 2021, 40(3): 61-69, 16.
IANG Y T, FENG W G. The focus of American think tanks on the “Silk Road of Health” initiative and China's countermeasures[J]. Journal of Intelligence, 2021, 40(3): 61-69, 16.

[4] 王飞 , 胡薇 . 中国 - 拉美共建“一带一路”的现状、问题与启示 : 基于智库研究视角 [J]. 重庆大学学报 (社会科学版), 2021, 27(4): 71-82.
WANG F, HU W. Realities, problems and reflections of China-Latin American cooperation under the Belt and Road initiative: A perspective from think tanks[J]. Journal of Chongqing University (Social Science Edition), 2021, 27(4): 71-82.

[5] 干保柱 , 李钰婧 . 日本各界及知名智库对“一带一路”倡议的关切和认知 [J]. 国外理论动态 , 2021(1): 98-103.
GAN B Z, LI Y J. Concerns and perceptions of all walks of life and well-known think tanks in Japan on the “the Belt and Road” initiative[J]. Foreign Theoretical Trends, 2021(1): 98-103.

[6] 陈媛 . 荷兰智库的“一带一路”观 : 海外话语实践建构 [J]. 国外社会科学 , 2021(1): 139-147, 160-161.
CHEN Y. How Dutch think tanks view the “Belt and Road”: Overseas discourse practice[J]. Social Sciences Abroad, 2021(1): 139-147, 160-161.

[7] VANGELI A. Diffusion of ideas in the era of the Belt and Road: Insights from China-CEE think tank cooperation[J]. Asia Europe Journal, 2019, 17(4): 421-436.

[8] 齐欣 . 基于推特的美国智库专家对华观察视角分析 : 以新美国安全中心为例 [J]. 情报理论与实践 , 2021, 44(7): 131-137.
QI X. Analysis of American think tank experts' observation perspective on China based on Twitter: Taking center for a new American security as an example[J]. Information Studies: Theory & Application, 2021, 44(7): 131-137.

[9] 齐欣 . 基于推特的美国智库专家涉华信息源分析 : 以新美国安全中心为例 [J]. 情报理论与实践 , 2020, 43(11): 104-109.
QI X. Analysis of China-related information sources of American think tank experts based on Twitter: Taking center for a new American security as an example[J]. Information Studies: Theory & Application, 2020, 43(11): 104-109.

[10] 王晰巍 , 刘宇桐 , 李玥琪 . 突发公共卫生事件中公民隐私泄露舆情的情感演化图谱研究 [J]. 情报理论与实践 , 2022, 45(3): 19-27.
WANG X W, LIU Y T, LI Y Q. A study on the emotional evolutionary mapping of public opinion on citizens' privacy leakage in public health emergencies[J]. Information Studies: Theory & Application, 2022, 45(3): 19-27.

[11] 孙瑞英 , 马晓伟 . 面向高校智库研究的多层次作者、主题与文本内容的演化分析 [J]. 情报科学 , 2020, 38(3): 167-176.
SUN R Y, MA X W. Evolution analysis of multi-level author, subject and textual content for university think tank[J]. Information Science, 2020, 38(3): 167-176.

[12] 王扶东 , 王苑潼 . 基于 LDA 模型的国内“一带一路”文献主题研究 [J]. 情报探索 , 2019(11): 129-134.
WANG F D, WANG Y T. Subject analysis of researches on the Belt and Road in China based on LDA model[J]. Information Research, 2019(11): 129-134.

[13] 胡潜 , 明均仁 . 基于用户 - 主题关联挖掘的虚拟社区推荐方法研究 [J]. 情报杂志 , 2017, 36(6): 156-159, 185.
HU Q, MING J R. Research on virtual community recommendation method based on user-topic association mining[J]. Journal of Intelligence, 2017, 36(6): 156-159, 185.

[14] 梁晓贺 , 田儒雅 , 吴蕾 , 等 . 基于超网络的微博舆情主

题挖掘方法 [J]. 情报理论与实践 , 2017, 40(10): 100-105.
LIANG X H, TIAN R Y, WU L, et al. A method of public opinion topic mining in micro-blog based on super-network[J]. Information Studies: Theory & Application, 2017, 40(10): 100-105.

[15] 曲靖野 , 陈震 , 胡轶楠 . 共词分析与 LDA 模型分析在文本主题挖掘中的比较研究 [J]. 情报科学 , 2018, 36(2): 18-23.
QU J Y, CHEN Z, HU Y N. A comparative study of co-word analysis and LDA model analysis in mining text topic[J]. Information Science, 2018, 36(2): 18-23.

[16] 沈思 , 李沁宇 , 叶媛 , 等 . 基于 TWE 模型的医学科技报告主题挖掘及演化分析研究 [J]. 数据分析与知识发现 , 2021, 5(3): 35-44.
SHEN S, LI Q Y, YE Y, et al. Topic mining and evolution analysis of medical sci-tech reports with TWE model[J]. Data Analysis and Knowledge Discovery, 2021, 5(3): 35-44.

作者贡献说明：
祁瑞华：论文理论框架与实验方法指导；
付 豪：实验开展与论文撰写。

Topic Mining and Evolution of “Belt and Road Initiative” Report in Think Tanks

Qi Ruihua[1,2]   Fu Hao[2]

[1] Language Intelligence Research Center of Dalian University of Foreign Languages, Dalian 116044

[2] School of Software of Dalian University of Foreign Languages, Dalian 116044

Abstract: [**Purpose/significance**] As an important part of the country’s soft power and voice, think tanks have a profound impact on government decision-making, enterprise development, public opinion and public knowledge dissemination. In terms of cognition of the “Belt and Road Initiative”, the research results of think tanks and think tank experts play an important role in foreign cooperation, public opinion guidance, project evaluation, etc., and can also reflect the work priorities of the “Belt and Road Initiative” and other information. Through quantitative research on the theme distribution and evolution of the research on the “Belt and Road Initiative” by Chinese think tanks, this paper obtains the key attention, research direction and research development context of the “Belt and Road Initiative” from the perspective of Chinese think tanks, and provides collation and reference for the promotion of relevant research on the “Belt and Road Initiative”. [**Method/process**] Using the LDA2vec theme method and calculation method of semantic similarity of word vector, we conduct theme mining and theme evolution research on 3,025 reports of Chinese think tanks listed in the *Global Think Tank Report 2020* from 2013 to 2020 on the Belt and Road, and carry out visual presentation to analyze the theme content and evolution characteristics. [**Result/conclusion**] The hot research topics on the “Belt and Road Initiative” by Chinese high-end think tanks cover political, economic, cultural, social, military, environmental and other fields year by year, and the research hot pots have a distinctive tone, such as “mutual benefit and mutual trust” “connectivity” “community of destiny” “win-win” and so on. The evolution of the Belt and Road Initiative’s theme reflects the characteristics of obvious theme inheritance and theme integration. This also reflects the gradual and comprehensive development process, and also shows that the research focus of domestic think tanks on the Belt and Road Initiative is sustainable and expanding.

Keywords: “Belt and Road Initiative”   topic mining   topic evolution

收稿日期：2022-01-20   修回日期：2022-03-14